



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 134765

TO: Jon E Angell  
Location: rem/2d20/2c18  
Art Unit: 1635  
Friday, October 08, 2004

Case Serial Number: 10/061201

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen 1A69  
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

Search Notes

O'Bryen, Barbara

---

**From:** Angell, Jon E  
**Sent:** Tuesday, October 05, 2004 3:35 PM  
**To:** O'Bryen, Barbara  
**Subject:** Sequence Database Search Request

SEARCH REQUEST FORM  
Scientific and Technical Information Center

Examiner# : 78697  
Art Unit : 1635  
Phone Number: 571-272-0756  
Date: 10/05/04  
Serial Number: 10/061,201  
Mailbox & Bldg/Room Location: REMSEN 2C18  
Results Format Preferred (circle): **DISK**

I would like to have a search of the standard databases performed using the following SEQ. ID NOs. from application :  
10/061,201

SEQ ID NO. 3

NOTE: SEQ ID NO: 3 is an amino acid sequence, but what I need is a search that identifies any NUCLEIC ACID SEQUENCE that encodes the amino acid sequence that is SEQ ID NO: 3.

Please search for the amino acid sequence of SEQ ID NO: 3 AND  
any nucleic acid that encodes the amino acid sequence of SEQ ID NO. 3

Thanks, Let me know if you need any further clarification...  
Eric

*J. Eric Angell*  
Art Unit 1635  
Office: REMSEN 2D20  
mailbox: REM 2C18  
571-272-0756

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 7, 2004, 08:28:49; Search time 6448 Seconds  
(without alignments)  
4900.290 Million cell updates/sec

Title: US-10-061-201-3

Perfect score: 3791

Sequence: 1 MDDLTLLDLEPCVCFEKL.....AKATTLVSTASGTQTVFPSK 729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delect 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=funet\_p2n model -DEV=xip  
-Q=cg2\_1USPTO\_spool\_pUS10061201/runat\_05102004\_174113\_9275/app\_query.fasta\_1\_903  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bis -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 -cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=prt -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10061201 @CGN 1.1 5600 @runat\_05102004\_174113\_9275 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOBJECT -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hig.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sis.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vt.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*

- 25: em\_pl.\*
- 26: em\_to.\*
- 27: em\_ss.\*
- 28: em\_un.\*
- 29: em\_vt.\*
- 30: em\_hig\_hum.\*
- 31: em\_hig\_inv.\*
- 32: em\_hig\_other.\*
- 33: em\_hig\_mus.\*
- 34: em\_hig\_pln.\*
- 35: em\_hig\_rdd.\*
- 36: em\_hig\_rnam.\*
- 37: em\_hig\_vt.\*
- 38: em\_sy.\*
- 39: em\_higo\_hum.\*
- 40: em\_higo\_mus.\*
- 41: em\_higo\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

% Query				Description
Result	No.	Score	Match Length DB ID	
1	3791	100.0	2190 6 AX530493	AX530493 Sequence
2	3791	100.0	2497 6 AX530492	AX530492 Sequence
3	3765	99.3	5608 9 HSM804610	AL833297 Homo sapi
4	3762	99.2	3019 9 AK074234	AK074234 Homo sapi
5	3219	84.9	1855 6 AX530497	AX530497 Sequence
6	3219	84.9	2086 6 AX530495	AX530495 Sequence
7	1446	38.1	1026 9 BC031650	BC031650 Homo sapi
8	1445	38.1	1008 9 AK058046	AK058046 Homo sapi
9	1167	30.8	2368 10 BC048400	BC048400 Mus muscu
10	1067	28.1	2685 10 AF515735	AF515735 Rattus no
11	1049	5.27	7667 6 AX776591	AX776591 Sequence
12	1049	5.27	5128 6 AX776593	AX776593 Sequence
13	1042	27.5	3206 6 AX776598	AX776598 Sequence
14	1042	27.5	3206 10 AF030131	AF030131 Mus muscu
15	1041	5.27	3251 6 AX540752	AX540752 Sequence
16	1033	5.27	2331 6 AX776594	AX776594 Sequence
17	1033	5.27	2332 6 AX878405	AX878405 Sequence
18	1033	5.27	2332 6 BD157200	BD157200 Primer fo
19	1033	5.27	2332 9 AK021429	AK021429 Homo sapi
20	1006	26.5	3146 6 AX675069	AX675069 Sequence
21	1004	26.5	750 6 AR413566	AR413566 Sequence
22	1004	26.5	750 6 BD109119	BD109119 EST and e
23	986	5.26	5668 9 BC053671	BC053671 Homo sapi
24	914	5.24	5074 10 BC060113	BC060113 Mus muscu
25	914	5.24	5074 10 BC060696	BC060696 Mus muscu
26	839	22.1	677 6 BD217010	BD217010 Novel hum
27	756	5.20	1977 3 AK112836	AK112836 Ciona int
28	756	19.9	5067 10 BC057304	BC057304 Mus muscu
29	745	5.19	1672 9 BC041023	BC041023 Homo sapi
30	708	5.18	3149 3 AY051984	AY051984 Drosophil
31	708	5.18	3149 6 AX776600	AX776600 Sequence
32	707	5.18	3123 3 AF220364	AF220364 Drosophil
c 33	677	17.9	195887 9 AC011359	AC011359 Homo sapi
34	675	17.8	500 6 AX530510	AX530510 Sequence
35	671	17.7	484 6 AX530500	AX530500 Sequence
36	666	5.17	612088 9 AC005216	AC005216 Homo sapi
37	655	5.17	317095 9 AC091887	AC091887 Homo sapi
38	626	16.5	534 6 AR413567	AR413567 Sequence
39	626	16.5	534 6 BD109120	BD109120 EST and e
40	609	5.16	1175506 3 AC008004	AC008004 Drosophil
41	609	5.16	1194897 3 AC007697	AC007697 Drosophil
42	609	5.16	1246095 3 AB003802	AB003802 Drosophil

43 604.5 15.9 177816 2 AC017643 AC017643 Drosophil  
44 601 15.9 359 6 AX530507 AX530507 Sequence  
45 596 15.7 500 6 AX530517 AX530517 Sequence

# ALIGNMENTS

## RESULT 1

LOCUS AX530493 2190 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 2 from Patent EP1239051.  
ACCESSION AX530493  
VERSION AX530493.1 GI:25252371  
KEYWORDS

## SOURCE

Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota, Metazoa, Chordata, Vertebrata, Euteleostomi,  
Mammalia, Eutheria, Primates, Catarrhini, Hominoidea, Homo.

## REFERENCE 1

AUTHORS Shannon,M.  
TITLE Human post-hic protein 1  
JOURNAL Patent: EP 1239051-A 2 11-SEP-2002,  
Aeomica, Inc. (US)

## FEATURES

Location/Qualifiers  
source 1..2190  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

## ORIGIN

Alignment Scores:  
Pred No.: 4,38e-149 Length: 2190  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-061-201-3 (1-729) x AX530493 (1-2190)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTTCACGTTACTTGTATCTTGGAGTGGCCCTGTGTCTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTCACAGCCAAAGTCTCCCTTGCACGACACCTTCTGCCAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAGAGCCCAAAAGAGCTGCGGTGCCCGCAATGACAGGACGCCTGTGTTTCCAAACATT 180  
Qy 61 GluAlaLeuProAlaAsnLeuLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGGGCGCTGCCGGCCAACTGCTGCTGCTGCGCCCTTCTGGATGGAGTGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlyGlySerArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGGAGAGGGGGCTCCCTCCGCGAGGCTGGCAGATGACCTTGCAGGATGGCAG 300  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGGACCAACCCACAGCTGTGCAGGCCAGTCTTTCGGGCTAGTGCCTAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGGTGCCTCGAGCAAAAGGCCCTTATGCCAACTACAGAGGGCAG 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu 160

Db 421 CCCGGTGACCTAAGGTTTAAATAAGGAGATATCATCTCTTCCCGGAGACAGCTTGATGAG 480  
Qy 161 AsnTrpTyrGlnGlyGluLeuAsnGlyLeuSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTGGTATCCAGGGGGAAATCAATGGCATCAGGGGAACTTCCAGCCAGCTCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCATCAAGCAGCTGCCCGCAGCCCGCCCGCTCTCGAGGGCCCTCTACAACTTGGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGlnAsnGlnAsnCysLysLeuThrPheLeuLysAspAspIle 220  
Db 601 GAGGCAAGGACAAAGAGTGAGAACACAGGATTGGCTGACCTTCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAAGGCAAGTTAGGAGATAAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuLysAsn 260  
Db 721 ATCTTCCCTATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTAGAGAAGAAC 780  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAGGTCCCGCAGTCATCTCTGCACAAAAAACCTGTCTCTGTCTCTCTCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgLysValProGlyGlnPheSer 300  
Db 841 AACAGTCTACCTCTCGTAGGGGCCCGGCTCCAGGAGGAAGTGGCTGGGCAGTTTTC 900  
Qy 301 IleThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAACAGCCTTTGAACACTCTCAACCCGATGGTGCCATTCTCTTCAGGGGCCCAT 960  
Qy 321 ValGluLeuSerThrProValLeuLeuSerSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATTCAGCACCCCGAGTGTCTACAGCTCCAGCAACCCCTCTGTGTATCACCCAG 1020  
Qy 341 MetGlnLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGAGAGCTTCTTCCAGCTGTGTGGGACAGGTCCAGCACTTATCACCC 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGACATTCCACAGCCGTGGTGTCTGTCTGCTGCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAAACATGTTTGTAGCCCTGCACTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGGAGAAAGGGCTCAGGGTCTCTGGGGAAGTGCCAGGACGGCTGCTCAGGG 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTCTGTGTCACCGGGGAGTCCGCACTTCCCAAAACAATTACGTATCCCATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAsnSerArgSerProGlyLeuTyrThrThrLeu 460  
Db 1321 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTGTCTACACCAATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerSerGlnGlySerIleSerGluGlyAspProArgGlnSerArg 480  
Db 1381 TCCACCTCTCTGTCTCTCCCAAGGCAGCATTTTCAGAAAGGTGATTCACGCAAAAGCC 1440  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500

Db 1441 CCCTTCAAATCCCGTCTTTGTGCCACTCCCATAGTCAACCCCGTGAGAGACACAGCCGGC 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerMetArgLysAsnGly 520  
Db 1501 CCTGGGACTTTAGGACAAAGGGTCTCTTCGGAAGGGGAGCAGCATGAGAAAGATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArg 540  
Db 1561 TCCCTGCAGAGACCCCTCCAGTCCGGGATCCCCCACTCTCTGTTGGTAGGCTCCCTCAGACGC 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheThrGlnProGlnGlyIlePro 560  
Db 1621 ACCCCACCAATGGTCTCTCCGCCCTCAGCAGTTCCAAATTTCTACCAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValGlnMetGlySerLysProAlaLeuThrGlyGlnPro 580  
Db 1681 TCCTCCCTCCCTCAGCCGTGGTGGAGATGGGGTCCAAAGCTGCCCTCAGCGGGAGCCCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGlnAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1741 GCCCTCAGTGCATCAGCAGGGGCAGTGAGGCCCGGACCCCACTCCGGGGCCAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProLysSerGluProLeuProLysProAlaSer 620  
Db 1801 ATTATGGAGACAAAGAAATCCCCATCAAGAGTGAAGCTCTGCGCAAAACCGCCGCATCT 1860  
Qy 621 AlaProSerSerLeuValLysProGluAsnSerArgAsnGlyIleGlnLysGlnVal 640  
Db 1861 GCCCCACCAATCACTCTGGTGAACACAGAAAACTCAAGAAATGGCATCGAAGCAAGATC 1920  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro 660  
Db 1921 AAAACCGTGAGATTTCAGAAATTACAGGCCCTCTCCACCAACAATTTACACCTCCCATCCC 1980  
Qy 661 ThrSerGlyLysProGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
Db 1981 ACCTCCGGAAGAGCTGAACAGCCAGCCACCCCTCAAGGGGTCCAGCCTGAAGCAGCGTCC 2040  
Qy 681 LeuGlyProGluMetThrValIlePheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2041 TTGGGGCCAGAGATGACCGTCTTATTTGGCCACCCGAAAGTGCGTGCCTCCTCCGACAGCAG 2100  
Qy 701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrThrLeuValSerThrAlaSer 720  
Db 2101 ACAGACCTCCGGAGAAAGTCAAGTCACTTGCCAAAGCCCAACCCCTGGTGTCTCACTGCCTCA 2160  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
Db 2161 GGCACGACGACCGTGTTCACGACAAA 2187

RESULT 2

AX530492

LOCUS AX530492 2497 bp DNA linear PAT 22-NOV-2002

DEFINITION Sequence 1 from Patent EP1239051.

ACCESSION AX530492

VERSION AX530492.1 GI:25252369

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota, Metazoa, Chordata, Craniata, Vertebrata, Euteleostomi,

Mammalia, Eutheria, Primates, Catarrhini, Homnidae, Homo.

REFERENCE 1

AUTHORS Shamon,M

TITLE Human posh-like protein 1

JOURNAL Patent: EP 1239051-A 1 11-SEP-2002,

Acomica, Inc. (US)

FEATURES Location/Qualifiers

source 1..2497

/organism="Homo sapiens"

/mol\_type="unassigned DNA"  
/db\_xref="taxon 9606"

ORIGIN

Alignment Scores  
Pred. No.: 509c-149 Length: 2497  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-061-201-3 (1-729) x AX530492 (1-2497)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 232 ATGGATGATTTGACGTTTACTTGATCTTTGGAGTGGCCCTGTGTGCTTTGAGAAGCTCGAT 291  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 292 GTCACAGCCAAAGTCTCCCTTGCCAGCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 351  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 352 TTCAAGGCCCCACAAAGAGCTGCGGTGCCCGGAAATGCAGGACGCGCTGTGTTTCCAAACATT 411  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 412 GAGCGCGTCCCGGCCAACCTGCTGCTGCTGGATGGAGTGGCGCTCAGGGCAG 471  
Qy 81 SerSerArgGlyGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 472 AGCTCCGGGAGAGGGGGCTCTTCCCGCAGGCGCTGGCAGCATGACCTTCGAGATGGCAGG 531  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 532 AAAAGCAGGACCAACCCAGACAGCTCTGCGGCCAGCTCTTCCCGGCTAGTGCTTAATGTC 591  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 592 AGAATCCACATGGATGGGTGGCTCGAGGCAAGGCCCTTATGCAACTACAGAGGGCAGAA 651  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 652 CCCGTCACCTAAGGTTTAAATAAGGGAGATATCATCTCTTCCGGAGACAGCTTGTATGAG 711  
Qy 161 AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 712 AATTGGTACCAGGGGGAAATCAATGGCATCAGCGGGAACCTTCCCGCCAGCTCCGTGGAA 771  
Qy 181 ValIleLysGlnLeuProProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 772 GTCATCAAGCAGCTGCCCCAGCGCCCGCTCTGCGAGGCGCTCTACAACTTCGACCTA 831  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220  
Db 832 CGAGGCAAGGACAAAGAGTGAGAACCCAGGATTGCGCTGACCTTCTCAAGGACGATATCATC 891  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 892 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCATCAGCGGGAACCTTCCCGCCAGCTCCGTGGAA 951  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 952 ATCTTCCTCTATCTTGTGTGTAGAGCCCAAAOCTCACCGCAAGACACCTTTTGTAGAGAA 1011  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 1012 AAAAGGTCCCGAGTCACTCTGCACAAAAAACCTGTCCCTGTGCTGCTCCTCCAGAGGC 1071

Qy 621 AlaProSerLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
Db 2092 CCCCCACCACTCCATCTGCTGGTGAACACGAGAAACTCAAGAAATGGGCATCGAAAGCAAGTC 2151  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro 660  
Db 2152 AAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACCAACAACATTACACCTCCCATCC 2211  
Qy 661 ThrSerGlyLysProGlnProAlaThrLysAlaSerGlnProGluAlaSer 680  
Db 2212 ACCCTCCGGAAGCGCTGAACAGCCAGCCACCCCTCAAGCGCTCCAGCGCTGAAGCAGCGTCC 2271  
Qy 681 LeuGlyProGluMetThrValLysPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2272 TTGGCCACAGAGATGACCGCTCTATTTCGCCACCGAAGTGGCTGCCACTCCGGACACGAG 2331  
Qy 701 ThrAspLeuArgArgLysSerAlaLysAlaLysAlaThrLysValSerThrAlaSer 720  
Db 2332 ACAGACCTCCCGAGAAAGATCAGCTCTTGCAAGGCCACCAACCTGGTGCCTGCTGCCTCA 2391  
Qy 721 GlyThrThrValPheProSerLys 729  
Db 2392 GCGACGCGACCGTGTTCGCCAGCAAA 2418  
RESULT 3  
LOCUS HSM804610 5608 bp mRNA linear PRI 13-MAY-2003  
DEFINITION Homo sapiens mRNA; cDNA DKFZp451J159 (from clone DKFZp451J159).  
ACCESSION AL833297  
VERSION AL833297.1 GI21733931  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 5608)  
AUTHORS Ottenwaelder B., Obermaier B., Deuschle B., Mewes H.W.,  
Weil B., Arnold C., Osanger A., Fobo G., Han M. and Wiemann S.  
TITLE Direct Submission  
JOURNAL Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr. 1, D-85764  
Neuberberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by MedGenomix (Martinsried/Germany) within the cDNA  
sequencing consortium of the German Genome Project. This clone  
(DKFZp451J159) is available at the RZPD in Berlin. Please contact  
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further  
information about the clone and the sequencing project is available  
at http://mips.gsf.de/proj/cDNA/  
FEATURES  
source 1..5608  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp451J159"  
/db\_xref="taxon:9606"  
/clone="DKFZp451J159"  
/tissue\_type="human skeletal muscle"  
/clone\_id="451 (synonym: hlec1) Vector pSport1; host  
DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
polyA\_signal 5564..5569  
polyA\_site 5588  
ORIGIN  
Alignment Scores: 1.54e-147 Length 5608  
Pred. No.: 3765.00 Matches: 725  
Score: 3765.00  
Percent Similarity: 99.45% Conservative: 0

Qy 231 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 1072 AACACGCTCTACCTCCGTAGGGGCCCCAGGGTCCAGGAGGAAGGTGGCTGGCGAGTTTTC 1131  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 1132 ATCAACAACAGCTTGAACACTCTCAACCGGATGGTCCATTCTCTTCAGGGCGCCATATG 1191  
Qy 321 ValGluIleSerThrProValLeuIleSerSerSerAsnProSerValIleThrGlnPro 340  
Db 1192 GTAGAGATCAGCACCCACCGCTCATCAGCTCCAGCAACCCCTCTGTGATCATCCAGCCCC 1251  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1252 ATGGAGAAAGCAGACGTTCTCTCCAGCTGTGGGACAGGTACAGCACTTATCACCCCGCA 1311  
Qy 361 ProValSerProGlyHisThrAlaValValSerLysProGlySerGlnGlnHisLeu 380  
Db 1312 CCTGTCTCTCCAGGACATTCACACGCCGTGGTCACTCTGCCCTCCAGCAACACCTC 1371  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1372 TCAGCGAACATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCGCATGAGCTGGAC 1431  
Qy 401 LeuGlnLysGlyGlnGlyValArgValLysGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1432 CTGCAAAAGGAGAGAGCGGTCAAGGCTCTGGGGAAAGTGCCAGGACGGCTGGCTCAGGGGC 1491  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1492 GTCTCTTGTACCGGGGAGTCGGCATCTTCCCAACAATTAAGTCATCCCATTTTC 1551  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrTrpThrLeu 460  
Db 1552 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCACATGGACGTTA 1611  
Qy 461 SerThrSerSerValSerSerGlnGlySerGlnGlyAspProArgGlnSerArg 480  
Db 1612 TCACCTCTCTGTGTCTCTCCCAAGCAGCATTCAGAAAGGTGATCCACGGCAAGCCGT 1671  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1672 CCTTCAAAATCCGTCTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAAGCACAGCCGGC 1731  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1732 CCTGGGACTTTAGGACAAAGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGATGGA 1791  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1792 TCCCTGCAGAGACCCCTCCAGTCCGGGATCCCACTCTCTGTGTAGGCTCCCTCAGACGC 1851  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1852 AGCCCCACCATGGTCTTCGGGCTCAGCAGTTCCTTACACAGCCACAGGGGATCCCC 1911  
Qy 561 SerSerProSerAlaValValGluMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1912 TCTCCCCCTCAGCGGTGGTGGAGATGGGGTCCAAAGCTGCCCTCAGGGGGAGCCCC 1971  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerSerLeu 600  
Db 1972 GCCCTCACGTGCATCAGCAGGGGCGAGTAGGGCCCGGACCCACTCCGGCGCCAGCTCCCTC 2031  
Qy 601 IleMetGluAspLysGluIleProlLysSerGluProLysProLysProAlaSer 620  
Db 2032 ATTAGAAGACAAAGAAATCCCCATCAAGAGTGGAGCTCTGCCCAAAACCCCGCATCT 2091

```

Best Local Similarity: 99.45%  Mismatches: 4
Query Match: 99.31%  Indels: 0
DB: 9  Gaps: 0

```

US-10-061-201-3 (1-729) x HSM804610 (1-5608)

Qy	1073	ATGGAGAAAGACAGACGTTCTTCACGCTGTGTGGGACAGGTGACGACTTATCACCCCGCA	1132
Db			
Qy	361	ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnGlnHisLeu	380
Db			
Qy	1133	CTGTCTCTCCAGGACATTCACAGCCGTTGGTCAGTCTGCTCGCTCCGACCAACACCTC	1192
Db			
Qy	381	SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp	400
Db			
Qy	1193	TCAGGGAAACAATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCGATGAGCTGGAC	1252
Db			
Qy	401	LeuGlnLysGlyGlnGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly	420
Db			
Qy	1253	CTGCAAAAGGGAGAGACGGCTCAGGGGTCTCGGGAAGTGCCTGGACGGCTGGCTCAGGGGC	1312
Db			
Qy	421	ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe	440
Db			
Qy	1313	GTCTCTGGTACCCGGCGGAGTCGGCATCTTCCCAACAATTACGTATCCCCATTTC	1372
Db			
Qy	441	ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu	460
Db			
Qy	1373	AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCACATGGACGTTA	1432
Db			
Qy	461	SerThrSerSerValSerGlnGlySerIleSerGlnGlyAspArgGlnSerArg	480
Db			
Qy	1433	TCCACCTCTCTGTCTCTCCCAAGGCACGATTTTCAGAAAGTGATCCACGGCAAGCCGT	1492
Db			
Qy	481	ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly	500
Db			
Qy	1493	CCCTCAAAATCGTCTTTGTGCCCACTGCCCATAGTCAACCCCGTGAGAACACACGCCGC	1552
Db			
Qy	501	ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly	520
Db			
Qy	1553	CTGGGACCTTTAGGACAAAGGGTGCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGAAATGA	1612
Db			
Qy	521	SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArg	540
Db			
Qy	1613	TCCTTCGAGAGACCCCTCCAGTCCGGGATCCCCACTCTCTGTGTAGGCTCCCTCAGACGC	1672
Db			
Qy	541	SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro	560
Db			
Qy	1673	AGCCCCACCATGGTCTTTCCGGCCCTCAGCATTTCCAAATTTCTACAGCCACAGGGGATCCCC	1732
Db			
Qy	561	SerSerProSerAlaValValValGlnMetGlySerLysProAlaLeuThrGlyGluPro	580
Db			
Qy	1733	TCCTCCCTCAGCCGTGGTGGAGATGGGTTCCAACGCTGCCTCCACGGGGAGGCC	1792
Db			
Qy	581	AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaIleSerSerLeu	600
Db			
Qy	1793	GGCTTCACGTGCATCAGCAGGGGCGATGAGGGCCGGATCCGCTCCCGGGCAGCTCCCTC	1852
Db			
Qy	601	IleMetGlnAspLysGluIleProIleLysSerGlnProLeuProLysProProAlaSer	620
Db			
Qy	1853	ATTATGGAAGACAAAGAAATCCCCATCAGAGTGAGCGCTCTGCAAAACCCGCCCATCT	1912
Db			
Qy	621	AlaProProSerIleLeuValLysProGlnAsnSerArgAsnGlyIleGluLysGlnVal	640
Db			
Qy	1913	GGCCCCACCATCCATCTGTGTGTAACACGAGAAACTCAAGAAATGGCATCGAAAGCAAGTC	1972
Db			
Qy	641	LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro	660
Db			
Qy	1973	AAACCGTGAGATTTCAGCAATTTACAGCCCTCTCCCAACCAACACTACACCTCCCATCC	2032

Qy 661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
D6 2033 ACCCCCGGAAAGCCTGAAACAGCCAGCCACCTCAAGGCGTCCAGCGCTGAAAGCAGCGTCC 2092  
Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
D6 2093 TTGGCCCGAGAGATGACCGTCTATTGGCCACCGAAGCGGCTGCCACTCCGAGACGAG 2152  
Qy 701 ThrAspLeuArgAlaLysSerAlaLeuAlaLysAlaThrThrLeuValSerThrAlaSer 720  
D6 2153 ACAGACCTCCCGAGAAAGTCAGCTCTTGCCAAAGGCCACCAACCCCTGGTGTCCACTCGCTCA 2212  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
D6 2213 GGCACCGCAGACCGTGTITCCCGAGCAAA 2239

RESULT 4  
AK074234  
LOCUS AK074234 3019 bp mRNA linear PRI 12-SEP-2003  
DEFINITION Homo sapiens cDNA FLJ23654 fis, clone COL10108.  
ACCESSION AK074234  
VERSION AK074234.1 GI:18676780  
KEYWORDS oligo capping, fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukarya, Metazoa, Chordata, Vertebrata, Euteleostomi,  
Mammalia, Eutheria, Primates, Catarrhini, Homnidae, Homo.

REFERENCE 1  
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,  
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,  
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3019)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.

TITLE Direct Submission  
JOURNAL Submitted (14-FEB-2002) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure, Human Genome  
Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan  
(E-mail: fldna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing;  
Research Association for Biotechnology; cDNA library construction,  
5'- & 3'-end one pass sequencing; Department of Virology and Human  
Genome Center, Institute of Medical Science, University of Tokyo  
(partly supported by Science and Technology Agency).

FEATURES Location/Qualifiers  
source 1..3019  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="COL10108"  
/issue\_type="colob"  
/clone\_lib="COL"  
/note="cloning vector pMEISFL3"  
223..1596  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB85025.1"  
/db\_xref="GI:18676781"

CDS  
/translation="MDDLTLDDLECPVCFBKLDVTAKVLPCQHTFCCKPCLQRVFAH  
KELRCEKTPVSNIEALPANLLVRLLDGVRSGSSGRGSRFRPTMTLQDGRKS  
RTNPRLOASPRLVNPHMDGVPRAKALCNVYQNGPDLRNRNGDILLRRLDE  
NWYGEISGNFNPASSVEYTKQLPQPPPLCRALYNFDLRGDKSENQDCLTFLKDD  
ITVISRVDENAEGRKDGKVGIFILFVFNLTARHLEKKNRGROSSCTKNLSVS  
SSRGNSTLRRGPGSRKVPVGGFSSITALTNLNRMVHPSGRHMYEISTPVLSSNP  
SVITQPMKADVPSSCGVQVSTYHPAPVPSGHSTA VVSLPGSQHLSANMFVALHSYS

AHGFDELIDLQKGEGRVRLGKCDQDQWLRGVS.LVTGRVGIFPNNNYVPIFRKTSPPDPSR  
SPGLYTT"

ORIGIN

Alignment Scores 101e-147 Length: 3019  
Pred. No.: 3762.00 Matches: 725  
Score: 3762.00  
Percent Similarity: 99.73% Conservative: 2  
Best Local Similarity: 99.45% Mismatches: 2  
Query Match: 99.24% Indels: 0  
DB: 9 Gaps: 0

US-10-061-201-3 (1-729) x AK074234 (1-3019)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
D6 223 ATGGATGATTTTGACGTTACTTTCTGGAGTGGCCCTGTGTGCTTTTGAGAAAGCTCGAT 282  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
D6 283 GTCACAGCCAAAGTCTCTCCCTTGCCAGCACACCTTCTGCAACCATGTCTACAGAGGGTT 342  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
D6 343 TTCAAGGCCCCACAAAGAGTGGCGGTGCCCGGAATGACGAGCGCTGTGTCTTTCACACATT 402  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
D6 403 GAGCGCTGCGCGCCAAACCTGTGCTGTGCGCCTTCTGGATGGAGTGGCGCTCAGGCGCAG 462  
Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
D6 463 AGCTCCGGGAGAGGGGGCTCTTCCGCGAGGCCCTGGCAGCATGACCTTTCAGGATGGCAGG 522  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerPheArgLeuValProAsnVal 120  
D6 523 AAAAGCAGGACCAACCCAGACGTCTGCGCCAGCTCTTCGCGCTAGTGCCTAATGTC 582  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
D6 583 AGAATCCACATGGATGGGTGCTCGAGCAAGAACGGCCCTTATGCAACTACAGAGGCGAGAAT 642  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
D6 643 CCCGGTGACCTAAGGTTTAAATAGGGGAGATATCATCCTTCTCCGGAGACAGCTTGTAG 702  
Qy 161 AsnTrpTyrGlnGlyGluLeuAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
D6 703 AATTGTGATCCAGGGGAAATCA GTGGCATCAGCGGAACTTCCAGCCAGCTCCGTGGAA 762  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
D6 763 GTCAATCAAGCAGCTGCCCCAGCGCCCGCTCTCGAGGGCCCTCTACAACCTTCGACCTA 822  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220  
D6 823 CGAGGCAAGGACAAAGAGGTGAGAAACCAAGGATTGGCTGACCTTCCCTAAGGACGATATCATC 882  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
D6 883 ACTGTGATCAGCCCGAGTGGATGAGAACTGGGCGAAGAGGCAAGTTTAGGAGATAAGTAGGC 942  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
D6 943 ATCTTCTCCTACTTCTTGTGTAGAGCCAAACCTCACCOCACAGACACCTTTTAGAGAAAGAAC 1002  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuLeuSerLeuValSerSerSerArgGly 280  
D6 1003 AAAAGGTGCCAGTCATCTCTGCACAAAAAAACCTGTCTCTCTGCTCTCTCCAGAGGC 1062



281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
1063 AACACAGTCTACCTCCGTAGGGGCCACAGGTCCAGGAGGAGGTGCTGGGCGATTTC 1122  
301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
1123 ATCAACAACAGCTTGAACACTCTCAACCGGATGGTCCATTCTCTTCAGGCGCCATATG 1182  
321 ValGluIleSerThrProValIleIleSerSerAsnProSerValIleThrGlnPro 340  
1183 GTAGAGATCAGCACCCCATGCTCATCAGCTCCAGCAACCCCTCTGTGATCAACCCAGCCC 1242  
341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
1243 ATGGAGAAAGCAGACGTTCTTCCAGCTGTGTGGACAGGTACAGCACTTATCAACCCGCCA 1302  
361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnHisLeu 380  
1303 CCTGTCTCTCCAGGACATTCCACAGCCGTGGTCACTCTGCCTGGCTCCACGAAACACCTC 1362  
381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
1363 TCAGCGAAACATGTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1422  
401 LeuGlnLysGlyGluGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
1423 CTGCAAAAGGGAGAGCGGTCAAGGTCTCTGGGGAAGTGCCAGGACGGCTGGGTCAAGGGC 1482  
421 ValSerLeuValThrGlyValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
1483 GTCTCTTGGTACCGGGGAGTCGGCATCTTCCCAAAACAATTACGTATCCCATTTTC 1542  
441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrTrpThrLeu 460  
1543 AGAAAGACCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACCACATGAACGTTA 1602  
461 SerThrSerValSerSerGlnGlySerIleSerGluGlyAspProArgGlnSerArg 480  
1603 TCCACCTCTCTGTGCTCTCCCAAGCAGCATTTCCAGAGGTGATCCACGGCAAGCCGT 1662  
481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
1663 CCCCCTCAAAATCCGTCTTTGTGCCCACTGCCCATAGTCAACCCCGTGAGAACAGCCGCGC 1722  
501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
1723 CTTGGGACTTTAGGACAAGGTCTCTTGGGAAAGGGCGGAGCAGCATGAGAAAGAAATGGA 1782  
521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
1783 TCCCTGCGAGAGACCCCTCCAGTCCGGGATCCCCACTCTCTGTGGTAGGCTCCCTCAGAGC 1842  
541 SerProThrMetValLeuArgProGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
1843 AGCCCAACCATGCTCTTGGGCTCAGCATGTTCCAAATCTACAGCCAGCGGGATCCCC 1902  
561 SerSerProSerAlaValValGluMetGlySerLysProAlaLeuThrGlyGluPro 580  
1903 TCTCCCTCAGCGGCTGGTGGTGGAGATGGGGTCCAAAGCTCGCCCTCAGGGGAGGCC 1962  
581 AlaLeuThrCysIleSerArgGlySerGlnAlaArgThrHisSerAlaAlaSerSerLeu 600  
1963 GCCCTCAGTGCATCAGCAGGGGCGAGTGAGGCCCGGATCCACTCGCGGCCAGCTCCCTC 2022  
601 IleMetGluAspLysGluIleProIleLysSerGluProLeuProLysProProAlaSer 620  
2023 ATTATGGAAGACAAAGAAATCCCCATCAAGAGTGAAGCTCTGTGCCAAAAACCGCCGCATCT 2082

621 AlaProProSerIleLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
2083 GCCCACCACCATCTCTGGTGAACACAGAAAACTCAAGAAATGGCATCGAAAAAGCAAGTC 2142  
641 LysThrValArgPheGlnAsnTyrSerProProProThrLysHisTyrThrSerHisPro 660  
2143 AAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACCAACCATTACACCTCCCATGCC 2202  
661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
2203 ACCTCCGGAAGAGCTGAACAGCCAGCCACCCCTCAAGGCGTCCAGCCTGAAAGCAGCGTCC 2262  
681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
2263 TTGGGCCAGAGATGACCGCTCTATTGCCCCACCGAAGTGGCTGCCACTCCGGACAGCAG 2322  
701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrLeuValSerThrAlaSer 720  
2323 ACAGACCTCCGGAGAAAGTCAAGCTCTTGCCAGGCCACACCCCTGGTGTCCACTGCCCTCA 2382  
721 GlyThrGlnThrValPheProSerLys 729  
2383 GGCACGCGACCGCTGTTTCCACGCAAA 2409

RESULT 5  
AX530497  
LOCUS AX530497 1855 bp DNA linear PAT 22-NOV-2002  
DEFINITION Sequence 6 from Patent EPI239051.  
ACCESSION AX530497  
VERSION AX530497.1 GI:25252377  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Shannon.M.  
TITLE Human post-like protein 1  
JOURNAL Patent: EP 1239051-A 6 11-SEP-2002;  
Aeomeica, Inc (US)  
FEATURES  
Location/Qualifiers  
source 1..1855  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,08e-125 Length: 1855  
Score: 3219.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.91% Indels: 0  
DB: 6 Gaps: 0  
US-10-061-201-3 (1-729) x AX530497 (1-1855)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Dy 1 ATGGATGATTTGACCGTTACTTGATCTTCTGGAGTGGCCCTGTGTGTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Dy 61 GTCAAGCCAAAGTCCCTCTTGGCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Dy 121 TTCAGGCCCCACAAAGAGTGGCGGTGCCCGGAATGCGAGGAGCGCTGTGTTTCCACATT 180

Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuValArgSerGlyGln 80  
Db 181 GAGGGCGCTGCCGGCAACCTGCTGCTGGCGCCCTCTGGATGGAGTGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGAGAGGGGGCTCTCCGACAGGCTGGCACGATGACCTTGCAGGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGACCAACCCACAGAGCTTCGAGGCCAGTCCCTTCGGGCTAGTGCTAATGTC 360  
Qy 121 ArgHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGTGCCCTCGAGCAAAAGGCCTTATGCAACTACAGAGGGCAGAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCCGGTGACCTAAGGTTAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480  
Qy 161 AsnTyrThrGlnGlyGluLeuAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATGGTACCAGGGGAAATCAATGGCATCAGCGGAACTTCCAGCCAGCTCGGTGGAA 540  
Qy 181 ValIleLysGlnLeuProProLeuLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCATCAAGCAGCTGCCCGCCAGCCGCTCTGCAGGGCCCTCTACAACTTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGlnAsnGlnAspCysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGGCAAGGCAAGAGTGAGAAACCAAGGATTGCCCTGACCTTCCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAGAAAGCAAGTTAGGAGATAAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 721 ATCTTCCTATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAGAAC 780  
Qy 261 LysGlyArgGlnSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAGGTGCCGCGATCTCTCGACAAAAAACTGTCCCTGGTGTCTCTCGCTCCAGAGGC 840  
Qy 281 AsnThrThrLeuArgArgGlyProGlySerArgLysValProGlyGlnPheSer 300  
Db 841 AACACGTCTACCTCCGTAGGGGCCAGGGTCCAGGAGGAAGGTGGCTGGGCGAGTTTCC 960  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAACAGCCTTTGAACACTCTCAACCCGGATGGTCCATCTCTCTCAGGGCGGCATATG 960  
Qy 321 ValGluIleSerThrProValIleSerSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAAGCAACCCAGTGTCTCATCAGCTCCAGCAACCCCTCTGTGATCATCCAGCCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGACGCTCTCTCCAGCTGTGTGGGACAGGTCAGCACTTATCACCCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCACACAGCCGTGGTCAGTCTGCCTGGCTCCACGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAAACATGTTTGTAGCCCTGCACCTCTACTCAGCCCAATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlnGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420

Db 1201 CTGCAAAAGGGAGAAAGGCGTCAGGGTCTCGGGAAAGTGCCAGGAGGCTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCCTTGGTCAACCGGGGAGTGGGCATCTTCCCAAAACAATACGTCAATCCCAATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu 460  
Db 1321 AGAAAGAGCTCTAGTTTTCAGACTCCCGGAGCCCTGGTCTCTACACACATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerSerGlnGlySerIleSerGlyAspProArgGlnSerArg 480  
Db 1381 TCCACCTCTCTGTCTCTCCCAAGGACGCAATTTTCAGAAAGGTGATCCACGGCAAGCCGT 1440  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1441 CCCTTCAAAATCCGCTCTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAGCACAGCCGCGC 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1501 CCTGGGACTTTAGGACCAAGGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGAAATGGA 1560  
Qy 521 SerLeuGlnArgProLysGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGCAGACACCCCTCCAGTCCGGGATCCCGGCTCTCTGTGTAGGCTCCTCTCAGACGC 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1621 AGCCCCACCATGGTCTCTTCGGCCCTCAGCAGTTCACAAITTTTACCAGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValGluMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1681 TCCTCCCTCAGCCGCTGGTGGTGGAGATGGGGTCCAGCGCTCCCTCAGGGGAGGCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaIleSerSerLeu 600  
Db 1741 GCCCTCAGCTGCATCAGCAGGGGCGAGTGAAGCCCGGACCCCACTCCCGCGGCCAAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIlePheLysSerGluProLeuProLysProPro 618  
Db 1801 ATTATGGAAGACAAAGAAATCCCCATCAGAGTGGCTCTGCCAAAAACCGCCC 1854

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: October 7, 2004 08:02:08 Search time 641 Seconds  
(without alignments)  
4831.414 Million cell updates/sec

Title: US-10-061-201-3

Perfect score: 3791

Sequence: 1 MDDLTLILLLECPVCFEKL.....AKATTLVSTASGTQTVFPSK 729

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 3373863 seqs. 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match: 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framer\_p2n\_model -DEV=exp  
-Q=lg2 /USPTO\_spool\_p/US10061201/runat\_05102004\_174113\_9264/app\_query.fasta\_1.903  
-DB=N Geneseq 29Jan04 -QFMT=fastp -SUFFIX=mg -MINMATCH=0 -1-LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blastm62 -TRANS=human40 cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=100 -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=epo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10061201 @CGN1\_1\_208 @runat\_05102004\_174113\_9264 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEO\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: N Geneseq 29Jan04.\*

- 1: geneseq1980s.\*
- 2: geneseq1990s.\*
- 3: geneseq2000s.\*
- 4: geneseq2001as.\*
- 5: geneseq2001bs.\*
- 6: geneseq2002s.\*
- 7: geneseq2003as.\*
- 8: geneseq2003bs.\*
- 9: geneseq2003cs.\*
- 10: geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	3791	100.0	2190	6	ABv89291 Human POS	
2	3791	100.0	2497	6	ABv89290 Human POS	
3	3219	84.9	1855	6	ABv89294 Human POS	
4	3219	84.9	2086	6	ABv89292 Human POS	
5	1408	5	37.2	1008	5	AAH64929 Human sec
6	1098	29.0	766	4	AAL01468 Human rep	
7	1098	29.0	766	4	ABH96913 Human tes	
8	1049	5	27.7	2667	8	ADB80966 RING-SH c
9	1049	5	27.7	5128	8	ADB80967 RING-SH c
10	1049	5	27.7	5205	7	ABX34585 Human mddl
11	1041	5	27.5	3251	6	ABN83326 Human cyt
12	1033	5	27.3	2331	8	ADB80968 RING-SH c
13	1033	5	27.3	2332	4	AAH15208 Human cDN
14	1033	5	27.3	2332	8	ADB80970 RING-SH c
15	1006	26.5	3146	7	ABv72517 Nucleotid	
16	839	22.1	677	2	AAz17679 Human gen	
17	816	5	21.5	1369	4	ABk3541 DNA enco
18	721	19.0	1701	6	ABk3562 Human cDN	
19	708	5	18.7	3122	4	ABL08347 Drocephal
20	675	17.8	500	6	ABv89306 Human POS	
21	671	17.7	484	6	ABv89296 Human POS	
22	651	17.2	793	5	AA569854 DNA enco	
c 23	609	5	16.1	5344	4	ABL08546 Drocephal
24	601	15.9	359	6	ABv89303 Human POS	

25	596	15.7	500	6	ABV89313 Human POS
26	595	15.7	359	4	AAI25317 Probe #15
27	595	15.7	359	4	ABA71252 Human fce
28	595	15.7	359	4	AAI51490 Probe #20
29	595	15.7	359	4	ABA37552 Probe #16
30	595	15.7	359	4	AAK45549 Human bon
31	595	15.7	359	4	AAK19553 Human bra
32	595	15.7	359	4	ABS45237 Human liv
33	595	15.7	359	6	ABS19820 Human sec
34	581	15.3	480	3	AAAC08830 Human sec
35	531	14.0	315	6	ABV89299 Human POS
36	528	13.9	500	6	ABV89309 Human POS
37	522	13.8	300	2	AAZ12879 Human gen
38	509	13.4	480	4	ABS33308 Human liv
39	509	13.4	480	6	ABS08393 Human gen
40	503	13.3	300	2	AAZ12878 Human gen
41	486	12.8	270	6	ABV89297 Human POS
42	486	12.8	500	6	ABV89307 Human POS
c 43	473	12.5	2241	5	AA572846 DNA enco
44	472	12.5	342	6	ABV89304 Human POS
45	472	12.5	500	6	ABV89314 Human POS

#### ALIGNMENTS

RESULT 1  
ABV89291  
ID ABV89291 standard: cDNA, 2190 BP.  
XX  
AC ABV89291;  
XX  
DT 23-DEC-2002 (first entry)  
XX  
DE Human POSHL1 encoding cDNA ORF SEQ ID NO 2.  
XX  
KW Human; POSHL1; SH3 domain; POSH-like signalling protein 1; oncogene;  
KW Rho GTPase; signal transduction; gene expression; cancer; vaccine;  
KW gene therapy; transgenic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2190  
FT /tag="a  
FT /product="POSHL1"  
XX  
PN EP1239051-A2.  
XX  
PD 11-SEP-2002  
XX  
PF 28-JAN-2002; 2002EP-00001165.  
XX  
PR 30-JAN-2001; 2001WO-US000663.  
PR 30-JAN-2001; 2001WO-US000664.  
PR 30-JAN-2001; 2001WO-US000665.  
PR 30-JAN-2001; 2001WO-US000666.  
PR 30-JAN-2001; 2001WO-US000667.  
PR 30-JAN-2001; 2001WO-US000668.  
PR 30-JAN-2001; 2001WO-US000669.  
PR 30-JAN-2001; 2001WO-US000670.  
PR 23-MAY-2001; 2001US-00864761.  
PR 10-OCT-2001; 2001US-0328205P.  
XX  
PA (AEOM)-AEOMICA INC.  
XX  
PI Shannon M.  
XX  
DR WPI 2002-684061/74  
DR P-PSDB; ABB83999.

XX PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.

XX PS Claim 1, Fig 3, 60pp + Sequence Listing, English.

XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (S1, ABB83999), a sequence having 65% sequence identity to (S1),  
CC (S1) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids.  
CC Human POSHL1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway. (I) is useful  
CC for identifying a specific binding partner. (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1, including diagnosing and  
CC treating cancer, they useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL1 cDNA sequence of the  
CC invention

XX SQ Sequence 2190 BP, 517 A, 701 C, 553 G, 419 T, 0 U, 0 Other;

Alignment Scores:

Pred No.: 1,7e-206 Length: 2190  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0

US-10-061-201-3 (1-729) x ABY89291 (1-2190)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTGACGTTACTTCTGGAGTGGCCCTGTGTGCTTTTGGAGAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTCCACAGCCAAAGTCTCCCTTGGCCAGCACACCTTCTGCCAAACCAATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAAAGGCCCAAAAGAGCTGCGGTGCCCGAATGCAAGACGCGCTGTGTTTTCACACATT 180  
Qy 61 GlnAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGGGCGCTCCCGGCCCAACCTGCTGCTGCTGGGCTTCTGGATGGAGTGGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlyGlySerArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGAGAGGGGGCTCCTTCCGCAGGCGCTGGCACATGACCTTGGCAGTGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAAGCAGGACCAACCCCAAGCTCTGCCAGGCCAGTCTGCCAGGCCAGTCTTCCGGCTAGTGCCTAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGTGCTCGAGCAAAAGGCCTTATGCCAACTACAGAGGGCAGAAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCCGGTGACCTAAGGTTAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480

Qy 161 AsnTrpTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTTGGTACCAAGCGGGAAATCAATATGGCATCAGCGGAACTTCCAGCAGCTCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCAATCAAGCAGCTGCCCGCCAGCGCGCCCGCTCTGCAGGGCCCTCTACAATCTGCACCTA 600  
Qy 201 ArgGlyLysAspLysSerGlnAsnGlnAspCysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGGCAAGGACAAGAGTGGAGAACAGGATTCGCTGACCTTCCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAGAAAGGCAAGTTAGGAGATAAGTAGG 720  
Qy 241 IlePheProIleLeuPheValGlnProAsnLeuThrAlaA'ghisLeuLeuGlnLysAsn 260  
Db 721 ATCTTCCCTACTCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAGAAC 780  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAAGTTCGCCAGTCACTCTGCACAAAAACCTGTCTCCCTGGTGTCTCGTCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgLysValProGlyGlnPheSer 300  
Db 841 AACACGCTCTACCTCTCGTAGGGGCCAGGGTCCAGAGGAAAGGTGCTGGCGCATTTTTC 900  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATCAACAACAGCCTTGAACACTCTCAACCGGATGGTCCCATTCCTCTCAGGGCGGCATATG 960  
Qy 321 ValGluIleThrProValLeuIleSerSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGCACCCCACTGCTCATCGCTCCAGCAACCCCTCTGTGTATCAACCCAGCCC 1020  
Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGACGCTTCTCTCCAGCTGTGTGGGACAGGTACAGCATTATCACCCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnHisLeu 380  
Db 1081 OCTGTCTCTCCAGGACATTCACACGCGGTGGTCACTGCTGCCGTGGCTCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAAACATGTTTGTAGCCCTGCACTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGAGAGAGCGGTCTGGGGAGTGGCAGGACGGCTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTTGTGTCACCGGGCGAGTCGGGATCTTCCCAAAACAATTACGTATCCCCATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrLeu 460  
Db 1321 AGAAAGACCTCTAGTTTTCACGACTCCCGGAGCCCTGGTCTCTACACCACATGGACGTTA 1380  
Qy 461 SerThrSerSerValSerSerGlnHisSerGluGlyAspProArgGlnSerArg 480  
Db 1381 TCCACCTCTCTGTGTCCTCCCAAGGCAGCATTTTCAGAAAGGTGATCCACGGCAAGCCGT 1440  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1441 CCCTTCAAAATCCGCTCTTTGTGCCACTGCCATAGTCAACCCCGGTGAGAAAGCACGCCGCG 1500  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520

Db 1501 CTGGGACTTTAGGCAAGGGTCTCTTCGGAAAGGGCGGAGCAGCATGAGAAAGATGGA 1560  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1561 TCCCTGACAGACCCCTCCAGTCGGGATCCCACTCTCTGGTGTAGGCTCCCTCAGAGCG 1620  
Qy 541 SerProThrMetValLeuArgProGlnGlnPheGlnPheTyrGlnProGlnGlyIlePro 560  
Db 1621 AGCCCCACCAATGGTCTCTGGGCTCAGCAGTTCACATTCACGCCACAGGGGATCCCC 1680  
Qy 561 SerSerProSerAlaValValValGlnMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1681 TCTCCCTCCCTCAGCCCGGTGGTGGAGATGGGGTCCAGCCCTGCCCCCACTCAGGGGGAGGCC 1740  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaAlaSerLeu 600  
Db 1741 GCCCTCAGCTGCATCAGCAGGGGAGTGAGGCCCGGACCCACTCCGCGGCCAGCTCCCTC 1800  
Qy 601 IleMetGluAspLysGluIleProLysSerGluProLysProLysProAlaSer 620  
Db 1801 ATTATGGAGAGACAAAGAAATCCCCATCAAGAGTGAAGCTCTGGCCAAAACCCCGGCATCT 1860  
Qy 621 AlaProProSerLeuValLysProGluAsnSerArgAsnGlyIleGluLysGlnVal 640  
Db 1861 GCCCCACCATCCATCTCTGGTGAAACACAGAAATCTCAAGAAATGGCATCGAAAGCAAGTC 1920  
Qy 641 LysThrValArgPheGlnAsnTyrSerProProThrThrLysHisTyrThrSerHisPro 660  
Db 1921 AAAACCGTGAGATTTCAGAAATACAGCCCTCTCCCAACCAACATTTACACCTCCCATCCC 1980  
Qy 661 ThrSerGlyLysProGluGlnProAlaThrLeuLysAlaSerGlnProGluAlaAlaSer 680  
Db 1981 ACCTCCGGAAGCCCTGAACAGCCAGCCAGCCACCTCAAGGGCTCCAGCGCTGACGAGCGTCC 2040  
Qy 681 LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln 700  
Db 2041 TTGGGGCCAGAGATGACCCGTCTATTTGCCCAACCGAAGTGCGCTGCCACTCCGGACAGCAG 2100  
Qy 701 ThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThrThrLeuValSerThrAlaSer 720  
Db 2101 ACAGACCTCCGGAGAAAGTCAGCTCTTGCACAGGCCACACCCCTGGTGTCTCCACTGCCTCA 2160  
Qy 721 GlyThrGlnThrValPheProSerLys 729  
Db 2161 GGCACGCAGACCGTGTTCACAGCAAA 2187  
RESULT 2  
ABV89290  
ID ABV89290 standard: cDNA, 2497 BP.  
XX AC ABV89290,  
XX AC ABV89290,  
DT 23-DEC-2002 (first entry)  
XX Human POSHL1 full length encoding cDNA SEQ ID NO 1.  
XX Human POSHL1: SH3 domain, POSHL-like signalling protein 1; oncogene,  
KW Rho GTPase; signal transduction, gene expression, cancer, vasculature,  
KW gene therapy; transgenic; gene; ss  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 232..2421  
FT /tag= a "POSHL1"  
FT /product= "POSHL1"  
XX  
PN BP1239031-A2.

XX  
PD 11-SEP-2002.  
XX  
PF 28-JAN-2002; 2002EP-00001165.  
XX  
PR 30-JAN-2001; 2001WO-US000663.  
PR 30-JAN-2001; 2001WO-US000664.  
PR 30-JAN-2001; 2001WO-US000665.  
PR 30-JAN-2001; 2001WO-US000666.  
PR 30-JAN-2001; 2001WO-US000667.  
PR 30-JAN-2001; 2001WO-US000668.  
PR 30-JAN-2001; 2001WO-US000669.  
PR 30-JAN-2001; 2001WO-US000670.  
PR 23-MAY-2001; 2001US-00864761.  
PR 10-OCT-2001; 2001US-0328205P.  
XX  
PA (AEOM-) AEOMICA INC.  
XX  
PI Shannon M.  
XX  
DR WPI; 2002-68406174.  
DR P-PSDB; ABB83999.  
XX  
PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL.  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.  
XX  
PS Claim 1; Fig 3; 60pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL 1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (SI, ABB83999), a sequence having 65% sequence identity to (SI),  
CC (SI) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids.  
CC Human POSHL 1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway. (I) is useful  
CC for identifying a specific binding partner (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1 including diagnosing and  
CC treating cancer, they are useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL 1 cDNA sequence of the  
CC invention  
XX  
SQ Sequence 2497 BP; 603 A; 777 C; 638 G; 479 T; 0 U; 0 Other;  
Alignment Scores:  
Pred No.: 199e-206 Length: 2497  
Score: 3791.00 Matches: 729  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 6 Gaps: 0  
US-10-061-201-3 (1-729) x ABV89290 (1-2497)  
Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 232 ATGGATGATTGACGTACTTCTCTGGAGTGGCCCTGTGTCTTTGAGAGGCTCGAT 291  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 292 GTCACAGCCAAAGTCTCTCCCTGCCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 351  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60

Qy	401	LeuGlnLysGlyGlnGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly	420
Db	1432	CTCAAAAGGGAGAAAGGGCGTCAAGGTCCTGGGGAAGTGCCAGGACGGCTGGCTCAGGGGC	1491
Qy	421	ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProllePhe	440
Db	1492	GTCTCTGGTACCCGGCGGAGTCGGCATCTTCCCAACAATACGTATCCCATTTTC	1551
Qy	441	ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu	460
Db	1552	AGAAAGACCTCTAGTHTTTCAGACTCCCGAGCCCTGGTCTCTACACCATGGACGTTA	1611
Qy	461	SerThrSerSerValSerGlnGlySerIleSerGlnGlyAspArgGlnSerArg	480
Db	1612	TCCACCTCTCTGTCTCCCAAGGCAGCATTTTCAGAAAGGTGATCCACGGCAAAAGCCGT	1671
Qy	481	ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly	500
Db	1672	CCCTTCAAAATCCGTCTTTGTGCCCACTGCCATAGTCAACCCCGTGAGAAAGCACAGCCGGC	1731
Qy	501	ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly	520
Db	1732	CCTGGGACTTTAGGACAAAGGGTCTTTCGGGAAAGGGCGGAGCAGCATGAGAAAGAATGGA	1791
Qy	521	SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg	540
Db	1792	TCCCTCAGAGACCCCTCCAGTCCGGGATCCCCACTCTCGTGGTAGCTCCCTCAGACGC	1851
Qy	541	SerProThrMetValLeuArgProGlnGlnPheThrGlnProGlnGlyIlePro	560
Db	1852	AGCCCCACCATGGTCTTTCGGCCCTCAGCATTCCTCAATTTACACGCCACAGGGGATCCCC	1911
Qy	561	SerSerProSerAlaValValGlnMetGlySerLysProAlaLeuThrGlyGluPro	580
Db	1912	TCTCCCCCTCAGCCGTGGTGGAGATGGGGTCCAAAGCTGCCCTCACGGGGAGGCC	1971
Qy	581	AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaIleSerLeu	600
Db	1972	GCCCTCACGTGCATCAGCAGGGGCGAGTGAAGGCCCGGAGCCCACTCCCGCGGCCAGCTCCCTC	2031
Qy	601	IleMetGlnAspLysGlnIleProlleLysSerGlnProLeuProLysProProAlaSer	620
Db	2032	ATTATGGAAGACAAAGAAATCCCCATCAAGATGAGGCTCTGCCAAACCGCCCGCATCT	2091
Qy	621	AlaProProSerIleLeuValLysProGlnAsnSerArgAsnGlyIleGlnLysVal	640
Db	2092	GCCGCCCACTCATCTCTGGTGAAACCAAGAAACTCAAGAAATGGCATCGAAAGCAAGTC2151	
Qy	641	LysThrValArgPheGlnAsnTyrSerProProThrLysHisTyrThrSerHisPro	660
Db	2152	AAAAACCGTGAGATTTTCAGAAATTACAGCCCTCTCCACAGCAAAACATTACAGCTCCCATCC	2211
Qy	661	ThrSerGlyLysProGlnProAlaThrLeuLysAlaSerGlnProGluAlaIleSer	680
Db	2212	ACCTTCGGAAAGCCTTGAAACAGCCACCCCTCAAGCGCTCCACGCCTGAAGCAGCGTCC	2271
Qy	681	LeuGlyProGluMetThrValLeuPheAlaHisArgSerGlyCysHisSerGlyGlnGln	700
Db	2272	TTCGGGCCAGAGATGCCGCTCTATTTCGCCCAACCGAAGTGGCTGCCACTCCGGACACGAC	2331
Qy	701	ThrAspLeuArgLysSerAlaLeuAlaLysAlaThrThrLeuValSerThrAlaSer	720
Db	2332	ACAGACCTCCGGAGAAAGTCAGCTCTTGCCAAAGGCCCAACACCTCGTGTCCACTGCTCA	2391
Qy	721	GlyThrGlnThrValPheProSerLys	729
Db	2392	GGCAGCAGACCGTGTTCCTCCAGCAA	2418

RESULT 3  
ABV89294  
ID ABV89294 standard; cDNA, 1855 BP.  
XX AC ABV89294.  
XX  
XX DT 23-DEC-2002 (first entry)  
XX  
XX DE Human POSHL1 partial encoding polynucleotide SEQ ID NO 6.  
XX  
XX KW Human; POSHL 1; SH3 domain, POSH-like signalling protein 1, oncogene;  
KW Rho GTPase; signal transduction, gene expression, cancer, vaccine,  
KW gene therapy, transgenic, ss.  
XX  
XX OS Homo sapiens.  
XX  
XX FN EPI239051-A2.  
XX  
XX PD 11-SEP-2002.  
XX  
XX PF 28-JAN-2002, 2002EP-00001165.  
XX  
XX PR 30-JAN-2001; 2001WO-US000663.  
XX PR 30-JAN-2001; 2001WO-US000664.  
XX PR 30-JAN-2001; 2001WO-US000665.  
XX PR 30-JAN-2001; 2001WO-US000666.  
XX PR 30-JAN-2001; 2001WO-US000667.  
XX PR 30-JAN-2001; 2001WO-US000668.  
XX PR 30-JAN-2001; 2001WO-US000669.  
XX PR 30-JAN-2001; 2001WO-US000670.  
XX PR 23-MAY-2001; 2001US-00864761.  
XX PR 10-OCT-2001; 2001US-0328205P.  
XX  
XX PA (AEOM-) AEOMICA INC.  
XX  
XX PI Shannon M.  
XX  
XX DR WPI, 2002-684061/74.  
XX  
XX PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.  
XX  
XX PS Example 1; SEQ ID NO 6; 60pp + Sequence Listing, English.  
XX  
XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL 1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (S1, ABB83599), a sequence having 65% sequence identity to (S1),  
CC (S1) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids  
CC Human POSHL 1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway (I) is useful  
CC for identifying a specific binding partner. (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1 including diagnosing and  
CC treating cancer; they are useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL 1 cDNA sequence of the  
CC invention. Note: The present sequence did not form part of the printed  
CC specification, but is based on sequence information supplied to Derwent  
CC by the European Patent Office  
XX  
XX SQ Sequence 1855 BP; 424 A; 581 C; 485 G; 365 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 4.7e-174 Length: 1855

Score: 3219.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.91% Indels: 0  
DB: 6 Gaps: 0  
US-10-061-201-3 (1-729) x ABV89294 (1-1855)  
Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTTTGACGTTACTTCTTGAGTGCCTCTGTGCTTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTCACAGCCAAAGTCTCTCCCTTGCCAGCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAAGGCCCCACAAAGAGCTGCGGTGCCCGGAATGCAGGACGCGCTGTGTTTCCAACATT 180  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGGCGCTGCCGGCCAACTGTCTGTGCGCCCTCTGGATGGAGTGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGGAGAGGGGCTCTTCCGACAGGCGCTGCGCAGCATGACCTTCAGGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGGACCAACCCAGAGCTCTCGAGGCCAGCTCTCCGGCTAGTGCCTAAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGTGCGCTCGAGGCAAGGCCCTTATGCAACTACAGAGGCGAGAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 421 CCCGTCACCTAAGGTTTAAATAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480  
Qy 161 AsnTyrTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 481 AATTGGTACCAGGGGAAATCAATGGCATCAGCGGGAATCTCCAGCCAGCTCCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCATCAAGCAGCTGCCCGCCAGCCCGCTCTCGAGGGCCCTCTACAACTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIleIle 220  
Db 601 CGAGGCAAGGACAAGAGTGAGAAACCAAGGATTCCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATCAAGCCGCGGTGAGAACTGGGCAAGAGGCAAGTTAGAGATAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuGluLysAsn 260  
Db 721 ATCTTCCTTATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAGAAC 780  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAAGGTGCGCAGTCATCTCTGCACAAAAAACCTGTCTCCTGTGTCCTCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 841 AACACGTCACCTCCGTAGGGGCCCAAGGTCAGGAGGAAAGGTGCTGGGCAGTTTCC 900  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320

DT 23-DEC-2002 (first entry)  
XX Human POSHL1 encoding cDNA nucleotides 1-2086 SEQ ID NO 4.  
XX  
XX Human: POSHL 1; SH3 domain; POSH-like signalling protein 1; oncogene;  
KW Rho GTPase; signal transduction; gene expression; cancer; vaccine;  
XX gene therapy; transgenic; ss  
XX Homo sapiens.  
XX  
XX PN EPI239051-A2.  
XX PD 11-SEP-2002.  
XX PF 28-JAN-2002; 2002EP-00001165.  
XX PR 30-JAN-2001; 2001WO-US0000663.  
XX PR 30-JAN-2001; 2001WO-US0000664.  
XX PR 30-JAN-2001; 2001WO-US0000665.  
XX PR 30-JAN-2001; 2001WO-US0000666.  
XX PR 30-JAN-2001; 2001WO-US0000667.  
XX PR 30-JAN-2001; 2001WO-US0000668.  
XX PR 30-JAN-2001; 2001WO-US0000669.  
XX PR 30-JAN-2001; 2001WO-US0000670.  
XX PR 23-MAY-2001; 2001US-00864761.  
XX PR 10-OCT-2001; 2001US-0328205P.  
XX PA (AEOM-) AEOMICA INC.  
XX  
XX PI Shannon M;  
XX  
XX DR WPI, 2002-684061/74.  
XX  
XX PT Novel human SH3 domain (POSH)-like signaling protein 1 polypeptide, POSHL  
PT -1, useful for treating disorders associated with decreased expression or  
PT activity of human POSHL1.  
XX  
XX PS Claim 4; SEQ ID NO 4; 60pp + Sequence Listing; English.  
XX  
XX CC The invention relates to an isolated SH3 domain (POSH)-like signalling  
CC protein 1 (POSHL 1) polypeptide (I), comprising a sequence of 730 amino  
CC acids (S1, ABB83999), a sequence having 65% sequence identity to (S1),  
CC (S1) having 95% deviations, especially conservative substitutions or a  
CC fragment of the sequences comprising at least 8 contiguous amino acids  
CC Human POSHL 1 is a proto-oncogene/oncogene product that functions as an  
CC adaptor protein that interacts with Rho family small GTPases as well as  
CC downstream components of the signal transduction pathway. (I) is useful  
CC for identifying a specific binding partner. (I) and nucleic acids (II)  
CC encoding (I) are useful for diagnosing, monitoring disease and treating  
CC caused by altered expression of human POSHL1 including diagnosing and  
CC treating cancer, they are useful in the development of vaccines and (II) is  
CC useful in gene therapy. (II) is useful for constructing microarrays which  
CC are useful for measuring and for surveying gene expression and creating  
CC transgenic non-human animals capable of producing the proteins. The  
CC present sequence is that of a human POSHL 1 cDNA sequence of the  
CC invention. Note: The present sequence did not form part of the printed  
CC specification, but is based on sequence information supplied to Derwent  
XX by the European Patent Office  
XX  
SQ Sequence 2086 BP; 482 A; 639 C; 558 G; 407 T; 0 U; 0 Other;  
  
Alignment Scores:  
Pred. No.: 539e-174 Length: 2086  
Score: 3219.00 Matches: 618  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 84.91% Indels: 0  
DB: 6 Gaps: 0  
  
RESULT 4  
ID ABV89292 standard; cDNA; 2086 BP.  
XX  
AC ABV89292;  
XX



US-10-061-201-3 (1-729) x ABV89292 (1-2086)

Qy 1 MetAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLeuLeuAsp 20  
Db 232 ATGGATGATTTGACGTTACTTGATCTCTGGAGTGCCTGTGTGTTGAGAAGCTCGAT 291  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 292 GTCAAGCCAAAGTCTCCCTTGCCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 351  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 352 TTCAAGGCCCAAAAGACCTGCGGTGCCCGCAATGCAGGAGCGCTGTGTTTCCAAACATT 411  
Qy 61 GluAlaLeuProAlaAsnLeuLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 412 GAGGGCGCTGCCGGGCCAACTGCTGCTGCTGCGCCTTCTGGATGGAGTGGCTCAGGGCAG 471  
Qy 81 SerSerGlyArgGlySerPheArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 472 AGCTCCGGGAGAGGGGGCTCTCCGCAAGGCTGGCAGCATGACCTTGCCAGGATGCGAGG 531  
Qy 101 LysSerThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 532 AAAAGCAGGACCAACCCAGAGCTCTGCAGGCCAGTCTTTCGGGCTAGTGCCTAATGTC 591  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLysLeuCysAsnTyrArgGlyGlnAsn 140  
Db 592 AGAATCCACATGGATGGGGTCCCTCGAGGCAAAAGGCCCTTATGCAACTACAGAGGCGAGAAT 651  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGlu 160  
Db 652 CCCGGTGACCTAAGGTTTAAATAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 711  
Qy 161 AsnTyrProGlnGlyIleSerGlyAsnPheProAlaSerSerValGlu 180  
Db 712 AATTGGTACACAGGGGGAATCAATGGCATCAGCGGGAACCTTCCAGCCAGCTCCGTGGAA 771  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 772 GTCATCAAGCAAGCTGCCCAAGCCCGCCCTCTGCAGGGCCCTCTACAACCTCGACCTA 831  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspIleIle 220  
Db 832 CGAGGCAAGGACAAAGAGTGAGAACCAAGGATTGCTGACCTTCTCAAGGACGATATCATC 891  
Qy 221 ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 892 ACTGTGATCAGCCGAGTGGATGAGAACTGGGCAGAAAGCAAGTTAGGAGATAAAGTAGGC 951  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuLeuLysAsn 260  
Db 952 ATCTTCCCTATCTTGTTGTAGAGCAAAACCTCACCGCAAGACACCTTTTAGAGAAGAAC 1011  
Qy 261 LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 1012 AAAAGTGGCCAGTCACTCTGTCACAAAAAACCTGTCCCTGTGCTGCTCTCCAGAGGC 1071  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 1072 AACAGTCTACCTCTAGGGGGCCAGGGTCCAGGAGGAAGGTGCTGCTGGGCACTTTTCC 1131  
Qy 301 IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 1132 ATCAACAACGCTTGAACACTCTCAACCGGATGGTGCCATTCTCTTCAGGGGGCCCATATG 1191  
Qy 321 ValGluIleSerThrProValLeuLeuSerSerAsnProSerValIleThrGlnPro 340  
Db 1192 GTAGAGATCAGCACCCAGTGCTCATCAGCTCCAGCAACCCCTCTGTGATCACCCAGCCC 1251

Qy 341 MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1252 ATGGAGAAAGCAGAGCTTCTCTCCAGCTGTGTGGACAGGTGAGCACTTATACACCCCGCA 1311  
Qy 361 ProValSerProGlyHisSerThrAlaValSerLeuProGlySerGlnHisLeu 380  
Db 1312 CCGTCTCTCCAGGACATTCCACACGCGTGGTCAGTCTGCTGCCGCCAGCAACACCTC 1371  
Qy 381 SerAlaAsnMetPheValAlaLysHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1372 TCAGCGAACATGTTTGTAGCCCTGCACTCTACTCAGCCCATGGACCCGATGAGCTGGAC 1431  
Qy 401 LeuGlnLysGlyGlyValArgValLeuGlyLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1432 CTGGCAAAAGGGAGAAAGGCGTCAAGGCTCTGGGAAAGTCCCAAGGCGCTGCTCAGGGGC 1491  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIlePhePhe 440  
Db 1492 GTCTCTCTTGGTCAACCGGGGAGTCGGCATCTTCCCAAAACAATTACGTCTACACCCATTTTC 1551  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu 460  
Db 1552 AGAAAGACCTCTAGTTTCCAGACTCCCGGAGCTGCCATAGTCAACCCGTGAGAGCACAGCCGGC 1611  
Qy 461 SerThrSerSerValSerSerGlnGlySerIleSerGlyAspProArgGlnSerArg 480  
Db 1612 TCCACCTCTCTGTGTCTCTCCCAAGGCAAGCATTTTCAGAAAGGTGATTCACCGCAAAAGCGGT 1671  
Qy 481 ProPheLysSerValPheValProThrAlaIleValAsnProValArgSerThrAlaGly 500  
Db 1672 CCCCTCAAAATCCGTTCTTGTGCCCACTGCCATAGTCAACCCGTGAGAGCACAGCCGGC 1731  
Qy 501 ProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArgSerSerMetArgLysAsnGly 520  
Db 1732 CCTGGGACTTTAGGACAAAGGGTCTCTTCGGAAAGGGGGAGCAGCATGAGAAAGATGGA 1791  
Qy 521 SerLeuGlnArgProLeuGlnSerGlyIleProThrLeuValValGlySerLeuArgArg 540  
Db 1792 TCCCTGCAGACACCCCTCCAGTCCGGGATCCCACTCTCGTGTAGGGTCCCTCAGACGC 1851  
Qy 541 SerProThrMetValLeuArgProGlnPheGlnPheThrTyrGlnProGlnGlyIlePro 560  
Db 1852 AGCCCCACCATGGTCTCTCGGCCCTCAGCAGTTCCTCAATTTCTACAGCCACAGGGGATCCCC 1911  
Qy 561 SerSerProSerAlaValValGlnMetGlySerLysProAlaLeuThrGlyGluPro 580  
Db 1912 TCTTCCCTCCCTCAGCCGTGTGTGGAGATGGGGTCCCAAGCTGCGCTCAGGGGGAGCCC 1971  
Qy 581 AlaLeuThrCysIleSerArgGlySerGluAlaArgThrHisSerAlaSerSerLeu 600  
Db 1972 GCCCTCAGCTGCATCAGCAGGGGGAGTGGAGCCCGGAGCCCACTCCGGGGGCCAGCTCCCTC 2031  
Qy 601 IleMetGluAspLysGluIlePheProLysSerGluProLysProPro 618  
Db 2032 ATTATGGAGACAAAGAAATCCCATCAAGAGTGGAGCTCTGCCAAAAACCCGCC 2085  
total number of his satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2u.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_pLUS10061201/nunat\_05102004\_174111\_9285/app\_query.fasta\_1\_903  
-DB=EST -QFMT=fastap -SUFFIX=rsi -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

UNITS=bits -START=1 -END=1 -MATRIX=blast62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pet -THR\_MIN=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ps -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10061201 @CGN 1 1 5180 @runat 05102004 174114 9285 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=5 -FGAPOP=6  
-FGAPOP=7 -YGAPOP=10 -YGAPOP=5 -DELOP=6 -DELEXT=7

Database : EST:

- 1: em\_estba.\*
- 2: em\_esthum.\*
- 3: em\_estin.\*
- 4: em\_estm.\*
- 5: em\_estov.\*
- 6: em\_estpl.\*
- 7: em\_estro.\*
- 8: em\_hic.\*
- 9: gb\_est1.\*
- 10: gb\_est2.\*
- 11: gb\_hic.\*
- 12: gb\_est3.\*
- 13: gb\_est4.\*
- 14: gb\_est5.\*
- 15: em\_estfun.\*
- 16: em\_estcom.\*
- 17: em\_gss\_hum.\*
- 18: em\_gss\_inv.\*
- 19: em\_gss\_pln.\*
- 20: em\_gss\_vit.\*
- 21: em\_gss\_fun.\*
- 22: em\_gss\_mam.\*
- 23: em\_gss\_mus.\*
- 24: em\_gss\_pro.\*
- 25: em\_gss\_rod.\*
- 26: em\_gss\_phg.\*
- 27: em\_gss\_vit.\*
- 28: gb\_gss1.\*
- 29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

%		Query		Match	Length	DB	ID	Description
Result	No.	Score	Match					
1	3615.5	95.4	2199	29	AY417552		AY417552	Homo sapi
2	3538.5	93.3	2199	29	AY417553		AY417553	Pan trogl
3	3088	81.5	3989	11	AK033601		AK033601	Mus muscu
4	2946.5	77.7	2205	29	AY417554		AY417554	Mus muscu
5	2891	76.3	4447	11	AK029168		AK029168	Mus muscu
6	2294	60.5	3182	11	AK028153		AK028153	Mus muscu
7	1454	38.4	865	13	BQ931816		BQ931816	AGENCOURT
8	1348	35.6	1114	12	BM453267		BM453267	AGENCOURT
9	1263	33.3	824	12	B1827373		B1827373	603077734
10	1240.5	32.7	931	12	B1519965		B1519965	603071753
11	1236	32.6	802	12	B1827187		B1827187	603077395
12	1236	32.6	972	12	B1520050		B1520050	603071361
13	1202	31.7	1158	10	BE615992		BE615992	601280026
14	1175	31.0	1159	12	B1825238		B1825238	603071926
15	1169.5	30.8	776	12	B1830329		B1830329	603073063
16	1167	30.8	819	10	BE736034		BE736034	601305637
17	1132.5	29.9	807	12	B1561960		B1561960	603254888
18	1124	29.6	772	12	BG218826		BG218826	RST38570
19	1056	27.9	721	12	B1461060		B1461060	603206926
c 20	1048	27.6	730	12	B1520952		B1520952	603071361
21	1046	27.6	660	10	BE616761		BE616761	601279705

22 1001.5 26.4 3521 11 AK029202  
23 993 26.2 691 10 BE311579  
24 982.5 25.9 867 13 BQ947482  
25 934 24.6 763 12 BG190764  
26 923 24.3 525 13 BX283463  
27 896.5 23.6 755 14 CA463900  
28 890 23.5 839 13 BU936713  
29 884 23.3 556 10 BU077029  
30 867 22.9 800 13 BU938619  
31 858.5 22.6 2036 11 BC033203  
32 856 22.6 769 14 CA466470  
33 852 22.5 593 13 BQ930446  
c 34 832 21.9 948 12 B1520780  
35 827.5 21.8 815 14 CA463777  
36 825 21.8 2406 11 AK088003  
37 816 21.5 683 14 CB854282  
38 816 21.5 1098 10 BB610623  
39 815 21.5 648 10 BB613665  
40 791 20.9 644 10 BB625180  
41 761 20.1 656 12 BM837901  
42 727 19.2 555 12 BM030927  
c 43 725 19.1 840 13 BU745401  
44 700 18.5 387 12 BM792505  
45 693 18.3 559 14 CB274053

ALIGNMENTS

RESULT 1  
AY417552  
LOCUS AY417552 2199 bp DNA linear GSS 12-DEC-2003  
DEFINITION Homo sapiens HCM6262 gene, VIRTUAL TRANSCRIPT, partial sequence,  
genomic survey sequence.  
ACCESSION AY417552  
VERSION AY417552.1 GI:39773512  
KEYWORDS GSS.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2199)  
AUTHORS Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous  
gene trios  
JOURNAL Science 302 (5652), 1960-1963 (2003)  
PUBMED 14671302  
REFERENCE 2 (bases 1 to 2199)  
AUTHORS Clark,A.G., Gianowski,S., Nielson,R., Thomas,P., Kejarwal,A.,  
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,  
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,  
Adams,M.D. and Cargill,M.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,  
Rockville, MD 20850, USA  
COMMENT This sequence was made by sequencing genomic exons and ordering  
them based on alignment.  
FEATURES Location/Qualifiers  
source 1..2199  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
gene <1..>2199  
/locus\_tag="HCM6262"  
ORIGIN  
Alignment Scores:

Prod. No.: 1.83e-255 Length: 2199

Score: 3616.50 Matches: 765

Percent Similarity: 93.92% Conservative: 1

Best Local Similarity: 95.79% Mismatches: 19

Query Match: 95.40% Indels: 11

DB: 25 Gaps: 2

US-10-061-201-3 (1-729) x AY417552 (1-2199)

Qy 1 MetAspAspLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGluLysLeuAsp 20  
Db 1 ATGGATGATTTCAGCTTACTTGAATCTTCGGAGTGCCTGTGTGCTTTGAGAAAGCTCGAT 60  
Qy 21 ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal 40  
Db 61 GTACACAGCCAAAGTCTCTCCGACAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT 120  
Qy 41 PheLysAlaHisLysGluLeuArgCysProGluCysArgThrProValPheSerAsnIle 60  
Db 121 TTCAGGGCCCCCAAGAGCTGCGGTGCCCGGAATGCAGGAGCGCTGTGTTTCCAACATT 180  
Qy 61 GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln 80  
Db 181 GAGGCGCTGCCGGCCAACTCTGCTGCTGCGGCTTCTGGATGGAGTGGCTCAGGGCAG 240  
Qy 81 SerSerGlyArgGlyGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg 100  
Db 241 AGCTCCGGGAGAGGGGGCTCTCCGACGGCTGGCACAGATGACCTTGCAGGATGGCAGG 300  
Qy 101 LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal 120  
Db 301 AAAAGCAGACCAACCCACGACGTCTGCAGGCCAGTCTTTCGGGCTAGTGCTAATGTC 360  
Qy 121 ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn 140  
Db 361 AGAATCCACATGGATGGGGTGCCTCGAGCAAGAGGCCTTATGCAACTACAGAGGCCAGAAT 420  
Qy 141 ProGlyAspLeuArgPheAsnLysGlyAspIleLeuLeuArgArgGlnLeuAspGly 160  
Db 421 CCCGGTGACCTAAGGTTTAAAGGGAGATATCATCTCTCCGGAGACAGCTTGATGAG 480  
Qy 161 AsnTyrThrGlnGlyGluLeuAsnGlyIleSerGlyAsnPheProAlaSerValGlu 180  
Db 481 AATTGGTACCAAGGGGGAATCAATGGGCATCAGCGGGAACCTTCCCAAGCACTCCGTGGAA 540  
Qy 181 ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu 200  
Db 541 GTCATCAAGCAGCTGCCAGCGCCCGCTCTGCAGGGCCCTCTACAACTTCGACCTA 600  
Qy 201 ArgGlyLysAspLysSerGluAsnGlnAspCysLeuThrPheLeuLysAspAspIle 220  
Db 601 CGAGGCAAGGACAAGGTGAGAACCAAGGATTGCGCTGACCTTCTCTCAAGGACGATATCATC 660  
Qy 221 ThrValIleSerArgValAspGluAsnTyrAlaGluGlyLysLeuGlyAspLysValGly 240  
Db 661 ACTGTGATACGCCGAGTGGATGAGAACTGGGCAGAGGCAAGTTAGGAGATAAAGTAGGC 720  
Qy 241 IlePheProIleLeuPheValGluProAsnLeuThrAlaArgHisLeuGluLysAsn 260  
Db 721 ATCTTCCTCATCTTGTGTAGAGCCAAACCTCACCGCAAGACACCTTTTAGAGAAAGAAC 780  
Qy 261 LysGlyArgGlnSerCysThrLysAsnLeuSerLeuValSerSerSerArgGly 280  
Db 781 AAAAGGTGCCAGTCAATCCCGCACAAAAAACCTGCTCCTGTCTGCTGCTCCTCCAGAGGC 840  
Qy 281 AsnThrSerThrLeuArgArgGlyProGlySerArgArgLysValProGlyGlnPheSer 300  
Db 841 AACACGTCTACCTCCGTAGGGGCCAGGGGTCCAGGAGGAAGGTGCTGGGCAAGTTTTC 900

Qy 301 IleThrThrAlaLeuAsnThrLeuLeuAsnArgMetValHisSerProSerGlyArgHisMet 320  
Db 901 ATACAACACAGCTTGAACACTCTCAACCGGATGGTGCCATTCTCTTCAGGGGCCCATATG 960  
Qy 321 ValGluIleSerThrProValLeuIleSerSerSerAsnProSerValIleThrGlnPro 340  
Db 961 GTAGAGATCAGCACCCAGTGCTCATCAAGCTCCAGCAACCCCTCTGTGTATCATCCAGGCC 1020  
Qy 341 MetGluLysAlaAspValProSerCysValGlyGlnValSerThrTyrHisProAla 360  
Db 1021 ATGGAGAAAGCAGAGCTTCTCTCCAGCTGTGGGACAGGTCAAGCACTTATACCCCCGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnHisLeu 380  
Db 1081 CCGTGTCTCTCAGGACATTCACAGCCGTGGTCAAGTCTGCTGCTCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAACATGTTGTAGCCCTGCCTACTCTACTCAGCCCATGGACCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGlyValArgValLeuGlyLysCysGlnAspGlyTyrLeuArgGly 420  
Db 1201 CTGCAAAAGGGAGAAAGGGGTCTCAGGGTCTGGGGAAGTGCCAGGAGCGCTGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIlePhePhe 440  
Db 1261 GTCTCTCTGTGTCACCCGGGCGAGTCGGGCACTTCCCAACAATTACGTCAATCCCAATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrThrThrLeu 460  
Db 1321 AGCACACCTCTCACAGTTATCCGGGTGTAGGGAGGTGGGCCA-----ACTTTA 1368  
Qy 461 SerThrSerValSer-----SerGlnGlySerIleSerGlu 473  
Db 1369 GTTAAAGCCTGTAGTTACAAAAATGGGTGTGTATGGGTGGGAAGAGGCAGCATTTTCAGAA 1428  
Qy 474 GlyAspProArgGlnSerArgProPheLysSerValPheValProThrAlaIleValAsn 493  
Db 1429 GGTGATCCACCGCAAAAGCGCTCCCTTCAATCCGCTCTTGTGCCCACTGCCCATAGTCAAC 1488  
Qy 494 ProValArgSerThrAlaGlyProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArg 513  
Db 1489 CCCGTGAGAGACACACCGCGCCCTGGGACTTTAGGACAAGGGTCTCTCGGAAAGGGCGG 1548  
Qy 514 SerSerMetArgLysAsnGlySerLeuGlnArgProLeuGlnSerGlyIleProThrLeu 533  
Db 1549 AGCAGCATGAGAAAGAAATGGATCCCTGCAGAGACCCCTCCAGTCCGGGATCCCACTCTC 1608  
Qy 534 ValValGlySerLeuArgSerProThrMetValLeuArgProGlnPheGlnPhe 553  
Db 1609 GTGGTAGGCTCCCTCAGACGAGCCGCCACCATGGTCTTTCGGCCTCAGCAGTTCCAATTTC 1668  
Qy 554 TyrGlnProGlnGlyIleProSerSerProSerAlaValValGluMetGlySerLys 573  
Db 1669 TACCAGCCACAGGGGATCCCTCTCCCTCCCTCAGCCGTGGTGGAGATGGGTCCAAG 1728  
Qy 574 ProAlaLeuThrGlyGluProAlaLeuThrCysIleSerArgGlySerGluAlaArgThr 593  
Db 1729 CCTGCCCTCAGGGGAGCCCGCCCTCACGTGCATCAGCAGGGGCGAGTGAGGCCCGGATC 1788  
Qy 594 HisSerAlaIleSerSerLeuIleMetGluAspLysGluIleProIleLysSerGluPro 613  
Db 1789 CACTCCCGGGCGGCGCTCCCTCATATTATGGGAAGACAAAGAAATCCCAATCAAGAGTGAGCCT 1848  
Qy 614 LeuProLysProProAlaSerArgAlaProProSerIleLeuValLysProGluAsnSerArg 633  
Db 1849 CTGCCAAACCCCGCCGCACTTGCCCAACCATCCATCTCTGTTGGTGAACCAAGAACTCAAGA 1908  
Qy 634 AsnGlyIleGluLysGlnValLysThrValArgPheGlnAsnTyrSerProProThrThr 653

Db	1909	AA1TGGCATCGAAAAAGCAAGTCAAAACCGTGAGATTTACAGAAATTCAGCCCTCTCCCAACC	1968			
Qy	654	LysHis1yThrSerHisProThrThrSerGlyLysProGluGlnProAlaThrLeuLysAla	673			
Db	1969	AAACAATTACACCTCCCATCCCACTCCGGGAAAGCCGTGAACACGCCAGCCACTCCCAAGGCG	2028			
Qy	674	SerGlnProGluAlaAlaSerLeuGlyProGluMerThrValLeuPheAlaHisArgSer	693			
Db	2029	TCCAGCTCGTGAAGACAGCGCTCTTGGGCCACAGAGATGACCGTGCTATTGGCCACCCGAAGT	2088			
Qy	694	GHyCysHisSerGlyGlnGlnThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThr	713			
Db	2089	GGCTGCCACTCCGGGACGACAGACAGCTCCGGAGAAAGTGACGCTCTGCGCAAGGCCACA	2148			
Qy	714	ThrLeuValSerThrAlaSerGlyThrGlnThrValPheProSerLys	729			
Db	2149	ACCTGGTGTCCTCGCTCAGGACGACAGACCGTGTTTCCACAGCAA	2196			
RESULT 2						
LOCUS	AY417553					
LENGTH	AY417553	2199 bp	DNA linear GSS 12-DEC-2003			
DEFINITION	Pan troglodytes HCM6262 gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.					
ACCESSION	AY417553					
VERSION	AY417553.1	GL39773513				
KEYWORDS	GSS					
SOURCE	Pan troglodytes (chimpanzee)					
ORGANISM	Pan troglodytes					
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.					
REFERENCE	1 (bases 1 to 2199)					
AUTHORS	Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarwal,A., Todd,M.A., Tenebaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.					
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trees					
JOURNAL	Science	302 (5652),	1960-1963 (2003)			
PUBMED	14571302					
REFERENCE	2 (bases 1 to 2199)					
AUTHORS	Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarwal,A., Todd,M.A., Tenebaum,D.M., Civello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J., Adams,M.D. and Cargill,M.					
TITLE	Direct Submission					
JOURNAL	Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA					
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.					
FEATURES	Location/Qualifiers					
source	1..2199					
	/organism="Pan troglodytes"					
	/db_type="genomic DNA"					
	/db_xref="taxon:9598"					
gene	<1..>2199					
	/locus_tag="HCM6262"					
ALIGN						
Alignment Scores:						
Pred No.:	1,02e-249	Length:	2199			
Score:	3538.50	Matches:	691			
Percent Similarity:	94.16%	Conservative:	2			
Best Local Similarity:	93.89%	Mismatches:	32			
Query Match:	93.34%	Indels:	11			
DB:	29	Gaps:	2			

Qy	1	MeAspAsnLeuThrLeuLeuAspLeuLeuGluCysProValCysPheGlnLysLeuAsp	20
Db	1	ATGGATGATTTCACGTACTTATGCTTCTTGAGATGCCCTGTGTGCTTTGGAGAAGCGTCGAT	60
Qy	21	ValThrAlaLysValLeuProCysGlnHisThrPheCysLysProCysLeuGlnArgVal	40
Db	61	GTCACGCCAAAGTCCTCCCTTGCACCAGCACACCTTCTGCAAAACCATGTCTACAGAGGGTT	120
Qy	41	PheLysAlaHisLysGlnLeuArgCysProGluCysArgThrProValPheSerAsnIle	60
Db	121	TTCAAGGCCCAACAAGAGCTGGGTGCCCGAATGCAGGACGCCTGTGTTTTCCAACATT	180
Qy	61	GluAlaLeuProAlaAsnLeuLeuValArgLeuLeuAspGlyValArgSerGlyGln	80
Db	181	GAGGCGCTGCCCGGCCAACCTTGCTGCTGTGCGCCTTCTGGATGGAGTGGCGCTCAGGGCAG	240
Qy	81	SerSerGlyArgGlyGlySerPheArgArgProGlyThrMetThrLeuGlnAspGlyArg	100
Db	241	AGCTCCGGGAGAGGGGGCTCCTTCCGACGCGCTGGCAGCATGACCTTCACAGATGGCAGG	300
Qy	101	LysSerArgThrAsnProArgLeuGlnAlaSerProPheArgLeuValProAsnVal	120
Db	301	AAAAGCAGGACCAACCCACGACGTTCGACGGCCAGCTCTTCGCGCTAGTGNNTAATGTC	360
Qy	121	ArgIleHisMetAspGlyValProArgAlaLysAlaLeuCysAsnTyrArgGlyGlnAsn	140
Db	361	AGAAATCCACATGGATGGGNNNNNGACAAANNNTTATGCAACTACAGAGGGCAGAAAT	420
Qy	141	ProGlyAspLeuArgPheAsnLysGlyAspIleIleLeuLeuArgArgGlnLeuAspGlu	160
Db	421	CCTGTGACCTTAAGGTTTAATAAGGGAGATATCATCTCTCTCCGGAGACAGCTTGATGAG	480
Qy	161	AsnTyrGlnGlyGluIleAsnGlyIleSerGlyAsnPheProAlaSerSerValGlu	180
Db	481	AATTTGGTACGAGGGGAAAACTAATGGCATCAGATGGGAACTTCCCACGACCTCGTGGAA	540
Qy	181	ValIleLysGlnLeuProGlnProProLeuCysArgAlaLeuTyrAsnPheAspLeu	200
Db	541	GTCATCAAAGCAGCTGCCCCACGCCGCCGCCCTCTCAGGGGCCCTCTACAACTTCGACCTA	600
Qy	201	ArgGlyLysAspLysSerGlnAsnGlnAspCysLeuThrPheLeuLysAspAspIle	220
Db	601	CGAGGCAAGGACACAAGATGAGTGAACCCAGGATTGCCTGACCTTCTCTCAAGGACGATATCATC	660
Qy	221	ThrValIleSerArgValAspGluAsnTrpAlaGluGlyLysLeuGlyAspLysValGly	240
Db	661	ACTGTGATCAGCCGAGTGGATGAGAACTGGGCGAGAAAGGCAAGTAGGAGATAAAGTAGGC	720
Qy	241	IlePheProIlePheValGluProAsnLeuThrAlaArgHisLeuLeuGlnLysAsn	260
Db	721	ATCTTCCCTATCTTGTTGTAGAGCAAAACCTCACCCAGACACCTTCTAGAGAAAGAAC	780
Qy	261	LysGlyArgGlnSerSerCysThrLysAsnLeuSerLeuValSerSerSerSerArgGly	280
Db	781	AAAAGTCCGACGTATCCCCGCAAAAAAACCTGTCTCTGGTGTCTCTGCTCTCCAGAGCG	840
Qy	281	AsnThrSerThrLeuLeuArgGlyProGlySerArgLysValProGlyGlnPheSer	300
Db	841	AACACGTCTACCTCCGTAGGGGGCCAGGGTCCAGGAGGAAGGTGCCTGGGCAATTTTTC	900
Qy	301	IleThrThrAlaLeuAsnThrLeuAsnArgMetValHisSerProSerGlyArgHisMet	320
Db	901	ATCACAACAGCCTTGAACTCTCAACCGGATGGTCCATTCTCTCTCAGGGCGCCCATG	960
Qy	321	ValGluIleSerThrProValLeuIleSerSerAsnProSerValIleThrGlnPro	340
Db	961	GTAGAGATCAGACCCCCAGTGCTCATCGTTCCAGCAACCCCTCTGTGATCACCCAGGCC	1020
Qy	341	MetGluLysAlaAspValProSerSerCysValGlyGlnValSerThrThrHisProAla	360

Db 2029 TCCACGGCTGAAGCAGCGTCTTGGGCCAGAGATGACCGTCTTATTTGGCCACCGAAGC 2088  
Qy 694 GlyCysHisSerGlyGlnThrAspLeuArgArgLysSerAlaLeuAlaLysAlaThr 713  
Db 2089 GGCTGCCACTCCGGACAGCAGACACTCCGGAGAAAAGTCAGCTCTTGGCAAGGCCACA 2148  
Qy 714 ThrLeuValSerThrAlaSerGlyThrGlnThrValPheProSerLys 729  
Db 2149 ACCCTGGTGTCCACTGCTCAGGCAACCGCAGACCGTGTCTTCCCAAGCAAA 2196

Db 1021 ATGGAGAAAGCAGACGGTCTCTCCAGCTGTGTGGGACAGGTCAAGCACTTATCACCCGGCA 1080  
Qy 361 ProValSerProGlyHisSerThrAlaValValSerLeuProGlySerGlnGlnHisLeu 380  
Db 1081 CCTGTCTCTCCAGGACATTCACAGCGGTGGTCACTGTCTGGCTGCCAGCAACACCTC 1140  
Qy 381 SerAlaAsnMetPheValAlaLeuHisSerTyrSerAlaHisGlyProAspGluLeuAsp 400  
Db 1141 TCAGCGAACATGTTTGTAGCCCTGCACCTCTACTCAGCCCATGGACCCCGATGAGCTGGAC 1200  
Qy 401 LeuGlnLysGlyGluGlyValArgValLeuLysCysGlnAspGlyTrpLeuArgGly 420  
Db 1201 CTGCAAAAGGGAGAAAGCGGTCAAGGTCTCTGGGGAAAGTGCACGAGCGGTGGCTCAGGGGC 1260  
Qy 421 ValSerLeuValThrGlyArgValGlyIlePheProAsnAsnTyrValIleProIlePhe 440  
Db 1261 GTCTCTTGGTCACCGGGCGAGTCGGCATCTTCCCAACAAATTACGTCACTCCCAATTTTC 1320  
Qy 441 ArgLysThrSerSerPheProAspSerArgSerProGlyLeuTyrThrTrpThrLeu 460  
Db 1321 AGCACACCCCTCACAGTTATCCGGGTGTAGGAGGTGGGGCCA-----ACTTTA 1368  
Qy 461 SerThrSerSerValSerSerGlnGly-----SerIleSerGlu 473  
Db 1369 GTTAAGCGCTGTAGNTCACAAAAATGGGTGTGTATGGGTGGGANNNNNGCAGCATTTTCAGAA 1428  
Qy 474 GlyAspProArgGlnSerArgProPheLysSerValPheValProThrAlaIleValAsn 493  
Db 1429 GGTGATCCACAGCAAAAGCGGTCCCTTCAATCCGTCTTTGTGCCCACTGCCATAGTCAAC 1488  
Qy 494 ProValArgSerThrAlaGlyProGlyThrLeuGlyGlnGlySerLeuArgLysGlyArg 513  
Db 1489 CCGTGAGAAAGCACAGCGGGCCCTGGGACTTTAGGACAAAGGTGNNNTGGGAAAGGCGG 1548  
Qy 514 SerSerMetArgLysAsnGlySerLeuGlnArgProLeuGlnSerGlyIleProThrLeu 533  
Db 1549 AGCAGCATGAGAAAGAAATGGATCCCTGCAGAGAACCCCTCCAGTCGGGGATCCCCACTCTC 1608  
Qy 534 ValValGlySerLeuArgSerProThrMetValLeuArgProGlnGlnPheGlnPhe 553  
Db 1609 GTGGTAGGCTCCCTCAGACGCAAGCCCAACCATGGTCTGCGGCTCAGCAGTTCCAATTTC 1668  
Qy 554 TyrGlnProGlnGlyIleProSerSerProSerAlaValValValGlnMetGlySerLys 573  
Db 1669 TACCGCCACAGGGGATCCCTCTCCCTCCCTCAGCCANNNNNGTGAGATGGGGTCCAAG 1728  
Qy 574 ProAlaLeuThrGlyGluProAlaLeuThrCysIleSerArgGlySerGluAlaArgThr 593  
Db 1729 CCTGCCCTCAGGGGGAGCCCGCCCTCAGGTGCATCAGCAGGGGCGAGGCGCCGGATC 1788  
Qy 594 HisSerAlaAlaSerSerLeuIleMetGluAspLysGluIleProIleLysSerGluPro 613  
Db 1789 CACTCCGGCGCCAGCTCCCTCATTTATGGAAAGACAAAGAAATCCCCCATCAAGAGTGAGCCT 1848  
Qy 614 LeuProLysProProAlaSerAlaProProSerIleLeuValLysProGluAsnSerArg 633  
Db 1849 CTGCCAAAACCCCGCCGCTCTGCCCAACCATCCATCTGTGTGAAACCCAGAAAACTCAAGN 1908  
Qy 634 AsnGlyIleGluLysGlnValLysThrValArgPheGlnAsnTyrSerProProThr 653  
Db 1909 AATGGCATCGAAAAANCAAGTCAAAAACCGGTGAGATTTCAGAAATTACAGCCCTCTCCACC 1968  
Qy 654 LysHisTyrThrSerHisProThrSerGlyLysProGluGlnProAlaThrLeuLysAla 673  
Db 1969 AAACATTACACCTCCCATCCCAACCTCTGGAAAGCCCTGAACAGCCAGCCACCCCTCAAGGCG 2028  
Qy 674 SerGlnProGluAlaSerLeuGlyProGluMetThrValLeuIleAlaHisArgSer 693